mation for a first set of objects, identifying additional information for a second set of objects, wherein the second set of objects is a subset of the first set of objects, classifying the second set of objects into multiple categories based on the additional information, training a first neural network and a second neural network based on the segmentation of the second set of objects, wherein the first neural network maps the object information to an embedding space and the second neural network maps the embedding space to a space of probability vectors corresponding to the multiple categories, mapping the first set of objects to the embedding space using the first neural network, classifying the first set of objects into the multiple categories using a low-density separation algorithm based on the mapping, retraining the first neural network and the second neural network based on the classification of the first set of objects, remapping the first set of objects to the embedding space using the updated first neural network, reclassifying the first set of objects into the multiple categories using the low-density separation algorithm based on the remapping, and retraining the first neural network and the second neural network based on the reclassification of the first set of objects.

[0127] Some examples of the method, apparatus, and non-transitory computer-readable medium described above may further include randomly selecting the second set of objects from the first set of objects. Some examples may further include collecting additional information based on the random selection. In some examples, the second set of objects is classified using unsupervised machine learning.

[0128] The description and drawings described herein represent example configurations and do not represent all the implementations within the scope of the claims. For example, the operations and steps may be rearranged, combined or otherwise modified. Also, structures and devices may be represented in the form of block diagrams to represent the relationship between components and avoid obscuring the described concepts. Similar components or features may have the same name but may have different reference numbers corresponding to different figures.

[0129] Some modifications to the disclosure may be readily apparent to those skilled in the art, and the principles defined herein may be applied to other variations without departing from the scope of the disclosure. Thus, the disclosure is not limited to the examples and designs described herein but is to be accorded the broadest scope consistent with the principles and novel features disclosed herein.

[0130] The described methods may be implemented or performed by devices that include a general-purpose processor, a digital signal processor (DSP), an application-specific integrated circuit (ASIC), a field-programmable gate array (FPGA) or other programmable logic devices, discrete gate or transistor logic, discrete hardware components, or any combination thereof. A general-purpose processor may be a microprocessor, a conventional processor, controller, microcontroller, or state machine. A processor may also be implemented as a combination of computing devices (e.g., a combination of a DSP and a microprocessor, multiple microprocessors, one or more microprocessors in conjunction with a DSP core, or any other such configuration). Thus, the functions described herein may be implemented in hardware or software and may be executed by a processor, firmware, or any combination thereof. If implemented in software executed by a processor, the functions may be stored in the form of instructions or code on a computer-readable medium.

[0131] Computer-readable media includes both non-transitory computer storage media and communication media including any medium that facilitates the transfer of code or data. A non-transitory storage medium may be any available medium that can be accessed by a computer. For example, non-transitory computer-readable media can comprise random access memory (RAM), read-only memory (ROM), electrically erasable programmable read-only memory (EEPROM), compact disk (CD) or other optical disk storage, magnetic disk storage, or any other non-transitory medium for carrying or storing data or code.

[0132] Also, connecting components may be properly termed as computer-readable media. For example, if code or data is transmitted from a website, server, or other remote source using a coaxial cable, fiber optic cable, twisted pair, digital subscriber line (DSL), or wireless technology such as infrared, radio, or microwave signals, then the coaxial cable, fiber optic cable, twisted pair, DSL, or wireless technology are included in the definition of medium. Combinations of media are also included within the scope of computer-readable media.

[0133] In this disclosure and the following claims, the word “or” indicates an inclusive list such that, for example, the list of X, Y. or Z means X or Y or Z or XY or XZ or YZ or XYZ. Also the phrase “based on” is not used to represent a closed set of conditions. For example, a step that is described as “based on condition A” may be based on both condition A and condition B. In other words, the phrase “based on” shall be construed to mean “based at least in part on.”

What is claimed is:

1. A method of user classification, comprising:

receiving user information for a first set of users;

receiving survey data for a second set of users, wherein the second set of users is a proper subset of the first set of users;

training a first neural network and a second neural network based on the second set of users, wherein the first neural network maps the user information to an embedding space and the second neural network maps the embedding space to a space of probability vectors, and wherein each vector in the space of probability vectors corresponds to a user’s category membership propensity;

mapping the user information for the first set of users to the embedding space using the first neural network;

predicting category membership propensities for the first set of users using a low-density separation algorithm on the user information for the first set of users mapped to the embedding space;

updating the first neural network and the second neural network based on the prediction; and

reclassifying the first set of users based on the updated first neural network and the updated second neural network.

2. The method of claim **1**, further comprising:

identifying demographic information for the first set of users; and